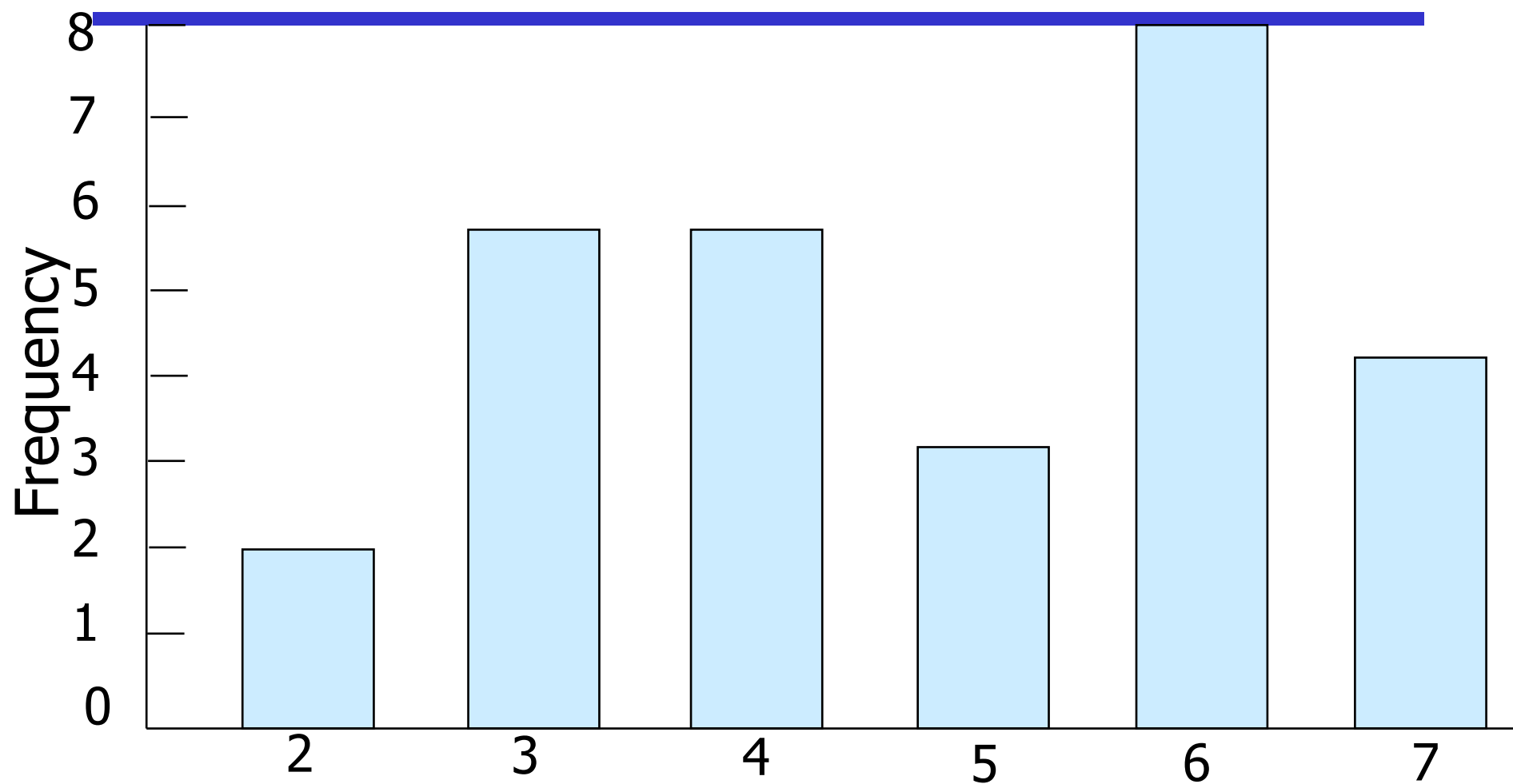




第十讲（第14章） 数据分析



频数直方图





集中趋势指标 (Measures of Location)

- 均值 *mean*

$$\bar{X} = \sum_{i=1}^n X_i / n$$

- 众数 **mode**

- 中位数 **median**



差异性指标 (Measures of Variability)

- 全距 **range**

$$\text{Range} = X_{\text{largest}} - X_{\text{smallest}}.$$

- 四分位差 **interquartile range**



差异性指标 (Measures of Variability)

- 方差 **variance**
- 标准差 **standard deviation**

$$s_x = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}}$$

- 变异系数 **coefficient of variation**

$$CV = s_x / \bar{X}$$



形状指标 (Measures of Shape)

- **Skewness** (偏度) .

$$S = \frac{E[(x - \bar{x})^3]}{\sigma^3}$$

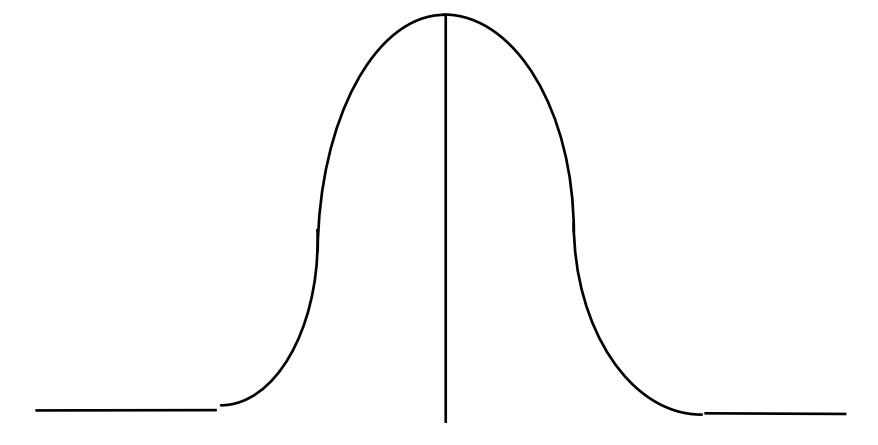
- **Kurtosis** (峰度)

$$K = \frac{E[(x - \bar{x})^4]}{\sigma^4} - 3$$



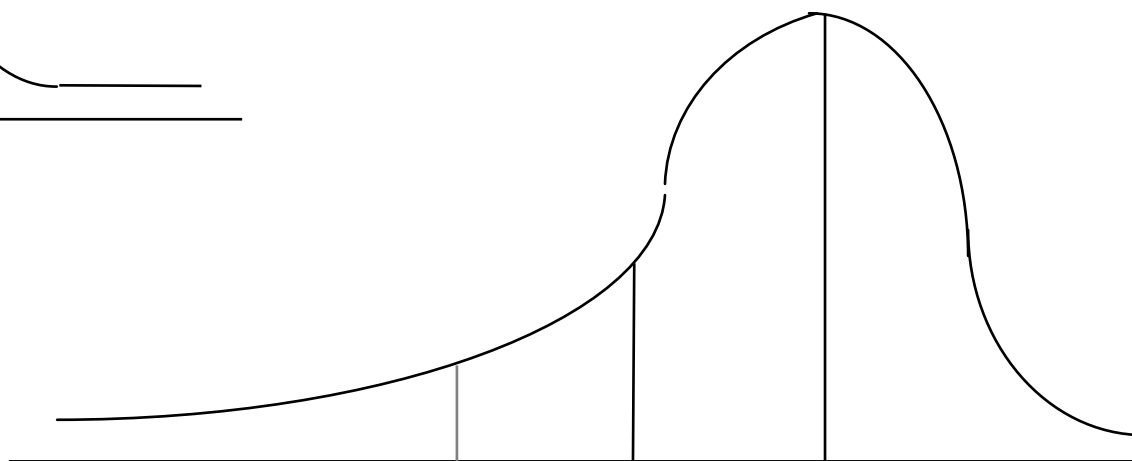
Skewness of a Distribution

Symmetric Distribution



Mean
Median
Mode
(a)

Skewed Distribution



Mean Median Mode
(b)



列联表

- 与频数分布不同，列联表（**cross-tabulation**）可同时描述两个或以上变量的情况。



性别和互联网使用频率

Gender

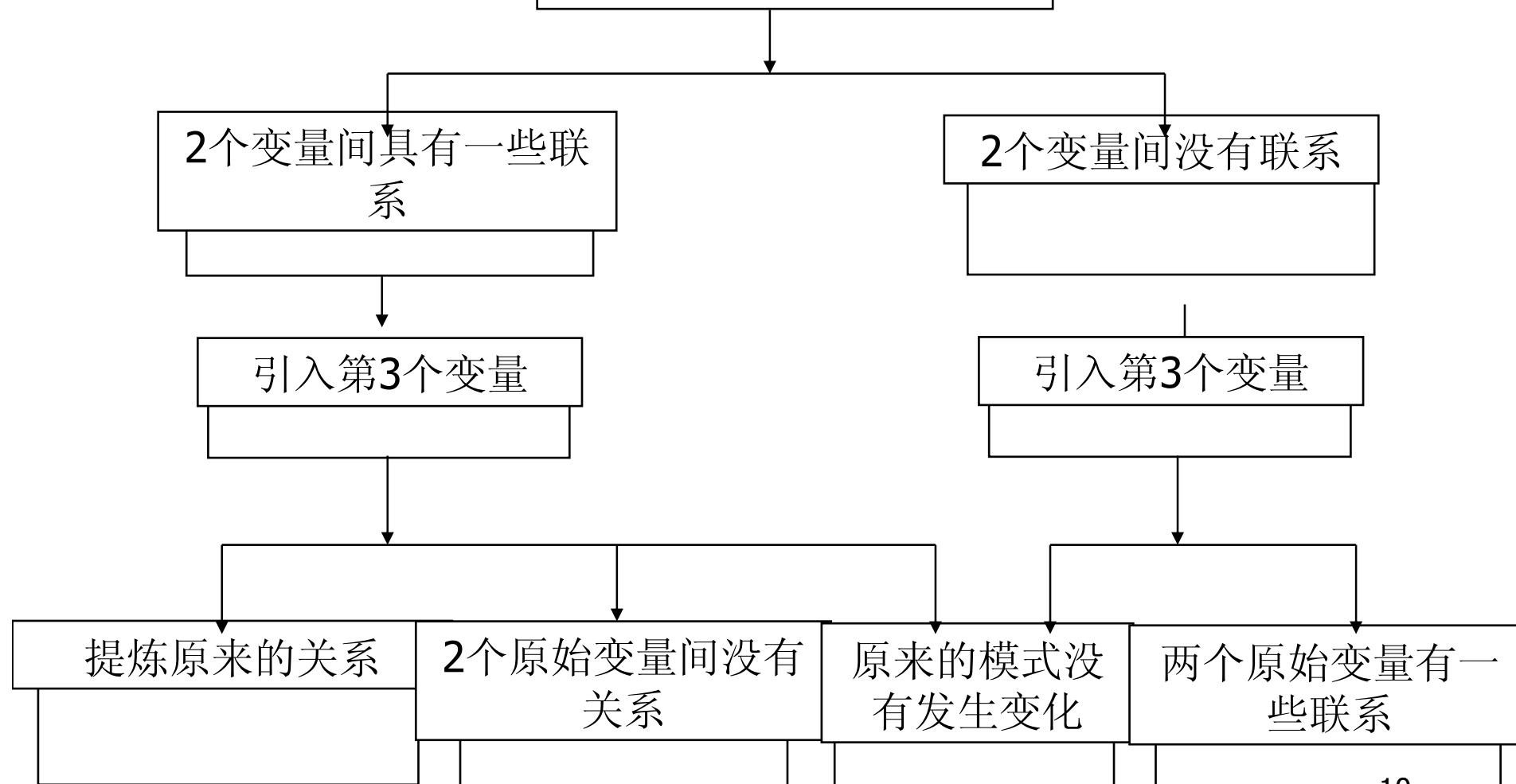
Internet Usage	Male	Female	Row Total
Light (1)	5	10	15
Heavy (2)	10	5	15
Column Total	15	15	





三变量交叉列表

两个原始变量





提炼原来的关系

Purchase of Fashion Clothing	Current Marital Status	
	Married	Unmarried
High	31%	52%
Low	69%	48%
Column	100%	100%
Number of respondents	700	300



Purchase of Fashion Clothing by Marital Status

Purchase of Fashion Clothing	Sex			
	Male		Female	
	Married	Not Married	Married	Not Married
High	<u>35%</u>	<u>40%</u>	<u>25%</u>	<u>60%</u>
Low	<u>65%</u>	<u>60%</u>	<u>75%</u>	<u>40%</u>
Column totals	100%	100%	100%	100%
Number of cases	400	120	300	180



原来的关系是虚假的

Own Expensive Automobile	Education	
	College Degree	No College Degree
Yes	32%	21%
No	68%	79%
Column totals	100%	100%
Number of cases	250	750



Ownership of Expensive Automobiles by Education Level and Income Levels

Own Expensive Automobile	Income			
	Low Income		High Income	
	College Degree	No College Degree	College Degree	No College Degree
Yes	20%	20%	40%	40%
No	80%	80%	60%	60%
Column totals	100%	100%	100%	100%
Number of respondents	100	700	150	50



揭示隐藏的联系

Desire to Travel Abroad	Age	
	Less than 45	45 or More
Yes	50%	50%
No	50%	50%
Column totals	100%	100%
Number of respondents	500	500



Desire to Travel Abroad by Age and Gender

Table 15.11

Desire to Travel Abroad	Sex			
	Male		Female	
	Age		Age	
	< 45	>=45	<45	>=45
Yes	60%	40%	35%	65%
No	40%	60%	65%	35%
Column totals	100%	100%	100%	100%
Number of Cases	300	300	200	200



原来的关系没有变化

Eat Frequently in Fast-Food Restaurants	Family Size	
	Small	Large
Yes	65%	65%
No	35%	35%
Column totals	100%	100%
Number of cases	500	500